Session 1

Building an efficient compression method for solar spectro-polarimetry data accumulated by Hinode/SP

J. Batmunkh[1], Y. Iida[1], T. Oba[2], H. Iijima[3]
[1] Niigata U., [2] NAOJ, [3] Nagoya U.

Solar spectra contain physical information that allows astronomers to broaden their knowledge about our sun. We have spectro-polarimetry data accumulated by the Hinode satellite for more than a decade. However, it is difficult to process such large amount of high dimensional data even with the present methods of informatics, including machine learning techniques. To this end, we aim to obtain a compressed representation of solar spectral data which is an important step for further detailed analysis of the polarimetric spectra, such as flare prediction, automatic categorization of spectra, detection of anomalous spectra, and so on.

We used Solar Optical Telescope-Spectropolarimeter (SOT-SP) data in this study. The spectro-polarimeter (SP) covers wavelength range between 630.1 and 630.3 nm, including Fe I line pair at 630.15 nm and 630.25 nm. The observation period of the data corresponds to 2021-08-03 and the field-of-view (FoV) of the 2D spatial spectro-polarimeter is 75.6″ ×81.2″ with a sampling slit of ~0.15″. The data was selected concerning its capture at near the disk center (not around the solar limb) and inclusion of both regular surface and sunspot regions.

We built an autoencoder, an encoder-decoder model based on deep learning techniques, for compressing Stokes I and V polarization parameters. The model gets polarization data as input and encodes and decodes it to output data which should be similar as possible to the input data. Output of the encoding part is our feature vector, namely the compressed representation of Stokes I and V parameters.

For the model training we constructed a customized loss function as the sum of mean absolute error (mae) of Stokes I and weighted mae of Stokes V. The weight value is determined as 0.1 which is the variance ratio of Stokes I and V continuum. We analysed the model performance from the correlation between raw and reconstructed spectra. Our study resulted in standard deviation (std) of 2.71~3.16% at the line centers and less than 0.7% at the continuum for Stokes I, and 4.46~4.79% at the line cores for Stokes V. The result shows that by using the customized loss function our model performed with smaller std values for reconstruction of the Stokes parameters.